This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ______

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

PATENT ABSTRACTS OF JAPAN

(11)Publication number : **09-097095**
(43)Date of publication of application : **08.04.1997**

(51)Int.CI. G10L 3/00
G10L 3/00
G06F 15/18

(21)Application number : **07-253146**
(22)Date of filing : **29.09.1995**

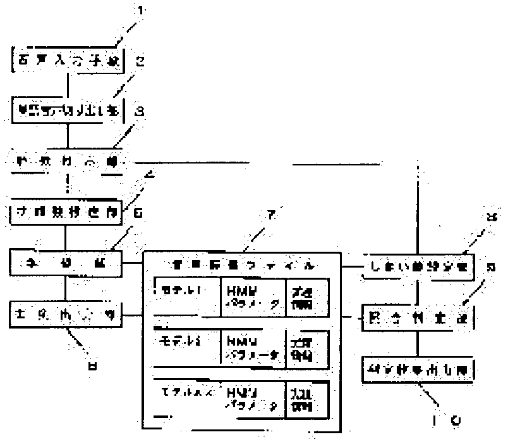
(71)Applicant : **MATSUSHITA ELECTRIC IND CO LTD**
(72)Inventor : **KONO NOBUYUKI**

(54) SPEECH RECOGNITION DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide the speech recognition device which uses HMM(Hidden Markov Model) that enables a rejection threshold for handling the input of an incorrect spoken sound such as a given cough to be set and used with likelihood corresponding to a user.

SOLUTION: In addition to a speech input means 1, a word voice segmentation part 2, a feature extraction part 3, a state estimation part 4, a learning part 5, etc., a likelihood output part 6 which finds likelihood from feature data and HMM parameters and a threshold setting part 8 which sets the threshold for rejection are constituted. Therefore, such trouble that a voice is rejected no matter how many times the user voices a word is eliminated and the usability of the user can be improved.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

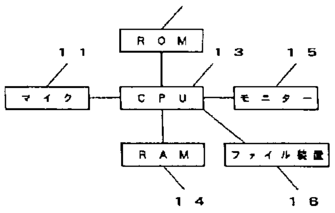
* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

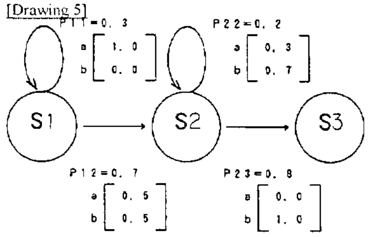
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

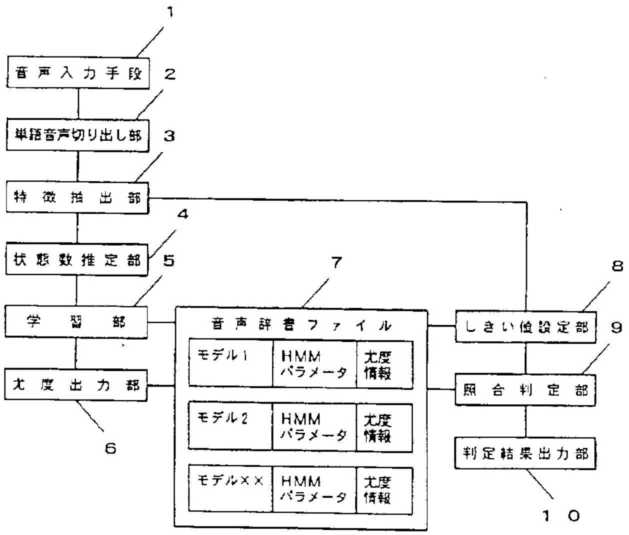
[Drawing 2]


1 2
ROM
1 1
マイク
CPU
1 3
1 5
モニター
RAM
ファイル装置
1 4
1 6

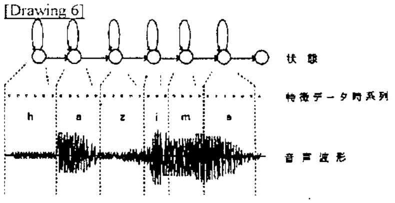
[Drawing 5]


P11=0. 3
a 1. 0
b 0. 0
P22=0. 2
a 0. 3
b 0. 7
S1
S2
S3
P12=0. 7
a 0. 5
b 0. 5
P23=0. 8
a 0. 0
b 1. 0

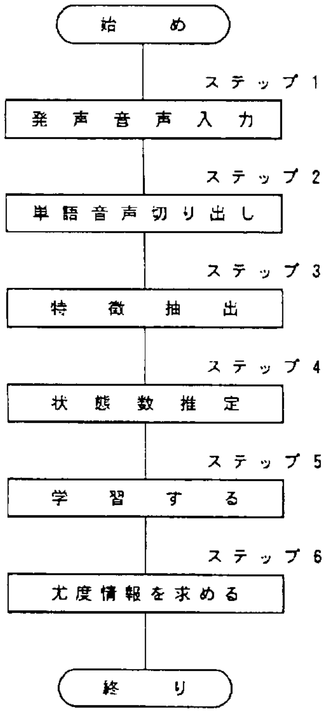
[Drawing 1]


1
音声入力手段
2
単語音声切り出し部
3
特徴抽出部
4
状態数推定部
5
学習部
尤度出力部
6
7
音声辞書ファイル
モデル1
HMM パラメータ
尤度情報
モデル2
HMM パラメータ
尤度情報
モデル××
HMM パラメータ
尤度情報
8
しきい値設定部
9
照合判定部
判定結果出力部
1 0

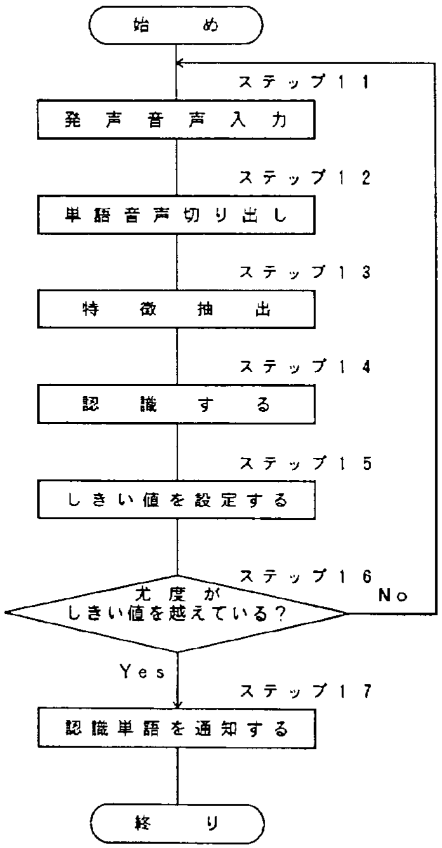
[Drawing 6]


状態
特徴データ時系列
h
a
z
i
m
e
音声波形

[Drawing 3]


始め
ステップ1
発声音声入力
ステップ2
単語音声切り出し
ステップ3
特徴抽出
ステップ4
状態数推定
ステップ5
学習する
ステップ6
尤度情報を求める
終り

[Drawing 4]


始め
ステップ11
発声音声入力
ステップ12
単語音声切り出し
ステップ13
特徴抽出
ステップ14
認識する
ステップ15
しきい値を設定する
ステップ16
尤度が
しきい値を越えている？
No
Yes
ステップ17
認識単語を通知する
終り

[Translation done.]

* NOTICES *

. Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] The configuration block Fig. of the voice recognition unit in one example of this invention
[Drawing 2] The circuit block diagram of the voice recognition unit in one example of this invention
[Drawing 3] The flow chart at the time of registration of the voice recognition unit in one example of this invention
[Drawing 4] The flow chart at the time of recognition of the voice recognition unit in one example of this invention
[Drawing 5] The conventional Hidden Markov Example Fig. of Model
[Drawing 6] The voice wave in the conventional speech recognition, the example Fig. showing the time series of the description data, and correspondence of each condition of HMM
[Description of Notations]
1 Voice Input Means
2 Word Voice Logging Section
3 Feature-Extraction Section
4 The Number Presumption Section of Conditions
5 Study Section
6 Likelihood Output Section
7 Voice Dictionary File
8 Threshold Setting Section
9 Collating Judging Section
10 Judgment Result Output Section
11 Microphone
12 ROM
13 CPU
14 RAM
15 Monitor
16 File Equipment

[Translation done.]

* NOTICES *

. Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] This invention recognizes word voice and relates to the voice recognition unit which outputs the recognition result.
[0002]
[Description of the Prior Art] The conventional Hidden Markov In order to explain the voice recognition unit which recognizes the word voice using Model (it is called HMM for short in this invention), the approach of the speech recognition by HMM is explained first. HMM is the Markov model of outputting one label (the description data) at a time by a certain probability (output probability) in that case while it has the conditions S1, S2, ..., SN of N individual and changes a condition one after another by a certain probability (transition probability) for every fixed period.
[0003] When voice is regarded as the time series of a label (the description data), HMM which uttered each word several times and modeled them at the time of study is created, and it recognizes by looking for HMM from which the probability (likelihood) which outputs the label sequence of input voice becomes max at the time of recognition. Hereafter, with reference to a drawing, it explains concretely.
[0004] Drawing 5 is the example Fig. of the conventional HMM, and is the easy example of HMM shown in Acoustical Society of Japan 42 Maki No. (1986) 12 "the speech recognition based on Hidden Markov Model." This HMM consists of three conditions and outputs the label sequence which consists only of two kinds of labels a and b. An initial state is S1 and changes from S1 to S1 the very thing by the probability of 0.3 (Label a is outputted in that case.). since a output probability is 0.0, Label b is outputted -- not having -- it changes to S2 by the probability of 0.7 (in that case, it is the probability of 0.5 about Label a, and Label b is outputted by the probability of 0.5). the probability of 0.2 from a condition S2 -- S2 the very thing -- changing (that time -- Labels a or b -- respectively -- 0. -- it outputting by the probability of 3 and 0.7) -- it changes to a final state S3 by the probability of 0.8 (Label b is outputted in that case.) Since a output probability is 0.0, it is outputted, and Label a is twisted and expresses things.
[0005] When the probability (likelihood) for this HMM to output the label sequence (train of the description data) abb is considered here, the condition sequence allowed by this HMM is only two, S1S1S2S3 and S1S2S2S3, and probabilities are 0.3*1.0*0.7*0.5*0.8*1.0=0.0840 and 0.7*0.5*0.2*0.7*0.8*1.0=0.0392, respectively. Since there is both of the possibility, it turns out by the probability of sum total 0.0840+0.0392=0.1232 that this HMM outputs abb.
[0006] Then, when asking for the transition probability in the condition of having been most suitable for each word, and the output probability of the label in each state transition and the label sequence of a certain strange word is inputted, if probability (likelihood) count is performed to each HMM, it will turn out HMM to which word tends to output this label sequence, and, thereby, recognition will be possible [that HMM is beforehand learned for every word, and]. The above is the approach of the speech recognition by HMM.
[0007] Moreover, drawing 6 is the example Fig. showing the time series of the voice wave in the conventional speech recognition, and the description data, and correspondence of each condition of HMM, and shows the correspondence the "start" and at the time of uttering. Thus, HMM is expressed in the condition with few number extent of phonemes of the word to the time series of the audio description data.
[0008] At the time of study, to each word registered into a voice recognition unit, it asked for the number of conditions with few number extent of phonemes of the word from spectrum change of a phoneme etc., the output probability of the description data in each state transition and the transition probability between conditions were presumed by study, and the model made in HMM, and input voice was applied to all these models, and it recognized by performing likelihood count in the voice recognition unit which recognizes the word voice using the conventional HMM at the time of recognition.
[0009]
[Problem(s) to be Solved by the Invention] Since it corresponds when there is an input of inaccurate utterance voice, such as a cough, in a voice recognition unit, it is important not to return a candidate with the always highest likelihood to a user, but to reject the candidate, if the candidate with the highest likelihood is not over a certain threshold, and to demand utterance from a user again, after that operability improves. However, since it is the constant value decided when the provider of a voice recognition unit evaluated that voice recognition unit beforehand, it will reject a recognition candidate and the threshold for this rejection may be unable to be recognized, even if it reuttered how many times for some users. Incidentally, according to the technical problem of the Chapter 10 10.2 speech recognition of the Furui ****** "digital signal processing" (Tokai University Press), "Although it is the speaker of few rates in the whole, there is a problem which a speaker with a very low recognition rate produces" is supposed.
[0010] By the speaker by whom a recognition rate becomes low, this It is because likelihood will be calculated lower than usual since it is out of range [the variation in an average speaker's description data] by the factor (for the purpose of being hard to happen probable). the difference in the plainness of voice with the description data peculiar to an individual -- mumbling -- etc. -- For this reason, the phenomenon in which the likelihood threshold set up based on an average speaker's description data always was not exceeded had occurred.

[0011] Therefore, this invention aims at offering the voice recognition unit using HMM which recognizes the word voice which makes it possible to set up and use the rejection threshold for corresponding to the input of inaccurate utterance voice, such as a cough, with the likelihood according to a user.

[0012]

[Means for Solving the Problem] For this reason, a voice input means for the voice recognition unit of this invention to input the voice containing word voice, The word voice logging section which starts only the part of voice to the word voice containing word voice, The feature-extraction section which extracts the description data from the started word voice, and the number presumption section of conditions which presumes the number of conditions to the word voice at the time of making a model by HMM from the description data, The study section which applies the description data to a word model and asks for a HMM parameter, The description data, the likelihood output section which asks for likelihood from a HMM parameter, and the voice dictionary file which consists of the HMM parameter and likelihood information which were learned, It had the threshold setting section which sets up the threshold for rejection, the collating judging section which calculates likelihood to each word model and judges a recognition candidate, and the judgment result output section which outputs a recognition result.

[0013]

[Function] The voice inputted into registration is recognized using the HMM parameter for which it learned and asked on the occasion o the word registration to a voice recognition unit, and it asks for the likelihood at that time. That is, even when it is the speaker to whom a recognition rate becomes low, it asks for the likelihood (it is lower) according to utterance of the speaker as a user's likelihood. And the likelihood is also registered into the voice dictionary file together with the HMM parameter. At the time of recognition, the likelihood information in a voice dictionary file is read, and it considers as the reference value of a likelihood threshold. Thereby, the likelihood threshold according to a user can be set up and exact recognition can be performed. Thus, since the likelihood threshold according to a user can be set up, "even if it reutters how many times for some users, it will reject a recognition candidate and cannot recognize" is lost.

[0014]

[Example] Hereafter, it explains, referring to a drawing about one example of this invention. Drawing 1 is the configuration block Fig. of the voice recognition unit in one example of this invention. A voice input means for one to input the voice containing word voice among drawing, the word voice logging section which starts only the part of voice to the word voice in which 2 contains word voice, The number presumption section of conditions which presumes the number of conditions to the word voice at the time of modeling the feature-extraction section which extracts the description data from the word voice which 3 started, and 4 by HMM from the description data, The study section which 5 applies the description data to a word model, and asks for a HMM parameter, The voice dictionary file which consists of the likelihood output section which 6 asks for likelihood from the description data and a HMM parameter, a HMM parameter which 7 learned, and likelihood information, The collating judging section which the threshold setting section to which 8 sets the threshold for rejection, and 9 calculate likelihood to each word model, and judges a recognition candidate, and 10 are the judgment result output sections which output a recognition result.

[0015] Drawing 2 is the circuit block diagram of the voice recognition unit in one example of this invention, and, for a read-only memory (ROM) and 13, as for the memory (RAM) which can be written in, and 15, a central processing unit (CPU) and 14 are [11 / a microphone and 12 / a monitor and 16] file equipment among drawing.

[0016] With a microphone 11, the voice input means 1 shown in drawing 1 the word voice logging section 2, the feature-extraction section 3, the number presumption section 4 of conditions, the study section 5, the likelihood output section 6, the threshold setting section 8, and the collating judging section 9 By performing the program memorized by ROM12 while CPU13 performed the exchange of a microphone 11, ROM12 and RAM14 and file equipment 16, and data The voice dictionary file 7 is realized by file equipment 16, and the judgment result output section 10 is realized by the monitor 15.

[0017] The flow chart at the time of registration of a voice recognition unit [in / in drawing 3 / one example of this invention] and drawing 4 are the flow charts at the time of recognition of the voice recognition unit in one example of this invention. The case where a certain word voice is registered into the voice recognition unit constituted as mentioned above is explained based on the flow chart of drawing 3 .

[0018] At step 1, the utterance voice containing word voice is inputted by the voice input means 1. At step 2, word voice is started from the utterance voice which contains word voice by the word voice logging section 2. This is [remove / a low noise part / silent or / before and behind word voice / detect and] realizable with audio power etc. In step 3, linear predictive coding (LPC analysis) performs a feature extraction in the feature-extraction section 3 by approaches, such as asking for the LPC cepstrum multiplier to the word voice. At step 4, the number of conditions to the word voice is presumed by the number presumption section 4 of conditions from the description data extracted from word voice at step 3. Presumption of the number of conditions can be performed based on the Acoustical Society of Japan lecture collected works (1990. 3) "the number of conditions and the number of mixing of HMM in continuation figure speech recognition."

[0019] At step 5, the HMM parameter for which learned using the HMM model with the number of conditions which asked for the description data of word voice at step 4 by the study section 5, and asked for the HMM parameter of the transition probability between each condition and the output probability of the description data in transition, and the voice dictionary file 7 was asked is stored. At step 6, likelihood count is performed on the HMM parameter for which it asked at step 5 read from the voice dictionary file 7 using the description data of word voice by the likelihood output section 6, and it asks for the likelihood. And the information on this likelihood is also stored in the voice dictionary file 7.

[0020] Next, this actuation is hereafter explained about the case where a certain word voice is recognized, based on the flow chart of drawing 4 . At step 11, the utterance voice containing word voice is inputted by the voice input means 1. At step 12, word voice is started from the utterance voice which contains word voice by the word voice logging section 2. At step 13, the feature-extraction section 3 performs the feature extraction to word voice. At step 14, likelihood count is performed on the HMM parameter of each word model read from the voice dictionary file 7 using the description data of word voice by the collating judging section 9, and a word model with high likelihood is judged to be a recognition candidate.

[0021] At step 15, the threshold for rejection is set up using the likelihood information read from the voice dictionary file 7 by the

threshold setting section 8. this threshold "makes a threshold read likelihood information as it is" -- it is made like or can set up as "let what carried out weighting for the likelihood information which read the voice recognition unit to the threshold evaluated and determined be a threshold" etc. At step 16, if the likelihood of the recognition candidate who asked at step 14 judged whether it would be over the threshold set up at step 15, and has exceeded by the collating judging section 9 and it has not progressed and exceeded to step 17, in order to reject and to have a user input again, it returns to step 11. At step 17, the judgment result output section 10 informs a user of a recognition result.
[0022]
[Effect of the Invention] As explained above, un-arranging [which it rejects even if a user utters how many times since the threshold of the rejection according to a user can be set up using the HMM parameter for which it learned and asked at the time of registration at the time of recognition by making the input voice at the time of registration recognize, and searching for the likelihood information according to the voice recognition unit of this invention] cannot arise, but a user's user-friendliness can be raised.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The word voice logging section which starts only the parts of voice to the voice input means for inputting the voice containing word voice, and the word voice containing word voice, The feature-extraction section which extracts the description data from the started word voice, and the number presumption section of conditions which presumes the number of conditions to the word voice at the time of making a model by HMM from the description data, The study section which applies the description data to a word model and asks for a HMM parameter, The description data, the likelihood output section which asks for likelihood from a HMM parameter, and the voice dictionary file which consists of the HMM parameter and likelihood information which were learned, The voice recognition unit characterized by having the threshold setting section which sets up the threshold for rejection, the collating judging section which calculates likelihood to each word model and judges a recognition candidate, and the judgment result output section which outputs a recognition result.

[Translation done.]